New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

A quantum Szilard engine without heat from a thermal reservoir

To cite this article: M Hamed Mohammady and Janet Anders 2017 *New J. Phys.* **19** 113026

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
14 June 2017

REVISED
18 August 2017

ACCEPTED FOR PUBLICATION
11 September 2017

PUBLISHED
16 November 2017

PAPER

A quantum Szilard engine without heat from a thermal reservoir

M Hamed Mohammady and Janet Anders

Department of Physics and Astronomy, University of Exeter, Stocker Road, Exeter, EX4 4QL, United Kingdom

E-mail: m.hamed.mohammady@gmail.com

Keywords: quantum thermodynamics, quantum measurement theory, Maxwell's demon

Abstract

We study a quantum Szilard engine that is not powered by heat drawn from a thermal reservoir, but rather by projective measurements. The engine is constituted of a system $\mathcal{S}$, a weight $\mathcal{W}$, and a Maxwell demon $\mathcal{D}$, and extracts work via measurement-assisted feedback control. By imposing natural constraints on the measurement and feedback processes, such as energy conservation and leaving the memory of the demon intact, we show that while the engine can function without heat from a thermal reservoir, it must give up at least one of the following features that are satisfied by a standard Szilard engine: (i) repeatability of measurements; (ii) invariant weight entropy; or (iii) positive work extraction for all measurement outcomes. This result is shown to be a consequence of the Wigner–Araki–Yanase theorem, which imposes restrictions on the observables that can be measured under additive conservation laws. This observation is a first-step towards developing 'second-law-like' relations for measurement-assisted feedback control beyond thermality.

1. Introduction

The possibility of extracting work from a system that is in thermal equilibrium, by means of measurement-assisted feedback control [1, 2], was first introduced by Maxwell [3, 4]. Seemingly violating the second law of thermodynamics, this observation sparked an intense debate, with a key contribution coming from Leo Szilard [5]. Szilard envisioned an engine where the system, $\mathcal{S}$, is a single particle in a box of volume V. Maxwell's demon, $\mathcal{D}$, extracts work from the system by performing two operations, namely, measurement and feedback. During the measurement stage, the demon places a frictionless partition inside the box, thus dividing it into two volumes V_L and V_R. Thereafter, the demon measures on which side the particle is located. During the feedback stage, conditional on the particle being found on the right (left) side of the partition, the demon attaches a weight-and-pulley mechanism to the right (left) of the partition so that, as the particle collides with the partition, the weight is elevated. The increase in the weight's gravitational potential energy is identified as the extracted work. This is shown schematically in figure 1.

By considering an infinite ensemble of such boxes, the average state of the particle can be interpreted as being an ideal gas occupying volume V_x for $x \in \{L, R\}$ which, after feedback, 'expands' to volume V. If the box is in thermal contact with a single reservoir $\mathcal{R}$ of temperature T, and the gas expands quasistatically, the engine will extract $W_x = K_B T \int_{V_x}^{V} \mathrm{d}V'/V' = K_B T \ln(V/V_x)$ units of work, where K_B is Boltzmann's constant. This is of course an average quantity of work, taken over the infinite ensemble of boxes. Moreover, the source of the extracted work is the heat drawn from the thermal reservoir. As the (average) state of the system at the start and end of the process is the same—an ideal gas occupying volume V—the Szilard engine is in apparent violation of the Kelvin statement of the second law; it is a cyclically operating device, the sole effect of which is to absorb energy in the form of heat from a single thermal reservoir and to produce an equal amount of work [6].

As shown by Penrose and Bennett [7–9], one may salvage the second law by observing that the demon is itself a physical entity, whose memory is altered by the measuring process. In order to make the engine cyclical the demon's memory must be returned to its initial configuration, i.e., the demon's memory must be 'reset' or 'erased'. If the erasure process is conducted by means of an interaction with the same thermal reservoir, it will

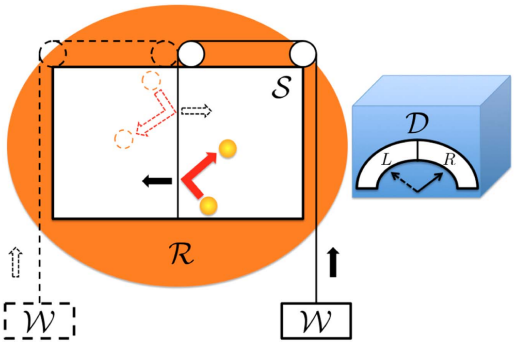

Figure 1. Szilard's engine. The demon, $\mathcal{D}$, places a partition inside a box containing a single particle. This is the system $\mathcal{S}$. During the measurement stage the demon measures the system and determines that the particle is on the right (left) hand side. This is stored in the demon's memory as the state R (L). During the feedback stage, the demon attaches a weight $\mathcal{W}$ to the partition via a pulley mechanism placed on the right (left) hand side. As the particle collides with the partition, moving it to the left (right), the weight is elevated and thus work is extracted. Each time the particle collides with the walls, it exchanges energy with the thermal reservoir, $\mathcal{R}$. As such, the source of work is the heat drawn from the reservoir.

require an average work cost no less than the average extracted work, which is dissipated as heat to the reservoir [10–12]; we may never win in the long run.

In recent years, much attention has been paid to the interplay between quantum theory and thermodynamics [13–23]. This has included the extension of work extraction through feedback control to the quantum regime, culminating in both theoretical [24–29] and experimental [30, 31] investigations. Of particular interest to our discussion is the work presented in [32, 33], wherein the authors consider the possibility of a Maxwell demon engine that functions in thermal isolation. Here, the source of work can no longer be identified as heat from a thermal reservoir, but rather as the energetic changes due to projective measurements. Such quantum measurements, however, ultimately result from a physical interaction between the system to be measured, and the measuring apparatus; in the case of a Szilard engine, the measuring apparatus is the demon's memory. It stands to reason, therefore, that energetic considerations come to bear on the measuring process [34–38], which will pose limitations on the performance of Szilard engines that, in lieu of a thermal reservoir, draw power from projective measurements.

We recall from the classical Szilard engine that hidden entropy sinks, when the demon's memory is not explicitly accounted for, allow for a violation of the second law. Similarly, hidden work sources involved in the measuring process can also allow us to 'cheat'. Consequently, a constraint of primary importance that must be imposed on the measuring process of a Szilard engine is energy conservation; if the energy of the system is increased by projective measurements, the demon's energy must decrease in kind. A central result from quantum measurement theory that is relevant to us is the Wigner–Araki–Yanase (WAY) theorem [39–44] which, under additive conservation laws, will limit the observables that can be measured. Using this, we shall show that while a Szilard engine can be powered by projective measurements instead of heat from a reservoir, it will have to give up at least one of three features that are present in the classical Szilard engine. The three features of the classical Szilard engine in question are:

Feature 1. The measurement is repeatable. If the demon measures the box and finds that the particle was on the right (left) hand side, a subsequent measurement would reveal that the particle is on the right (left) hand side with certainty. This allows for the interpretation that, after the measurement has been completed, the system 'possesses' the revealed value.

Feature 2. The weight's entropy does not change as a result of work extraction. Work is extracted by raising the weight, thus increasing its gravitational potential energy. In general, the height of the weight's center of mass will be a fluctuating quantity, with an uncertainty Δh. However, Δh does not change as a result of work extraction. In other words, the weight is neither 'cooled' nor 'heated' as it is elevated.

Feature 3. The engine works reliably—the work extracted is strictly positive for all measurement outcomes. Whether the particle is on the right or left hand side of the box, the extracted work has the value $W_x = K_B T \ln(V/V_x)$ where $x \in \{L, R\}$. As V and $V_x < V$ are always positive, finite numbers, then $W_x > 0$ for all $x \in \{L, R\}$.

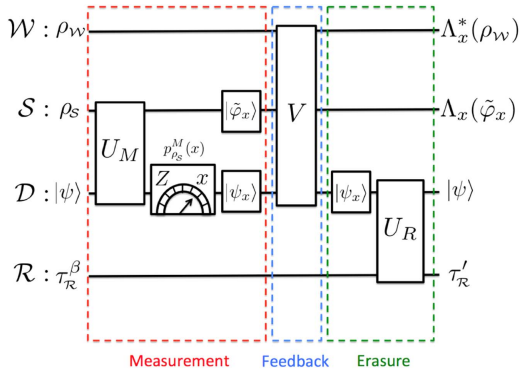

Figure 2. The circuit model of measurement-assisted work extraction, without heat from the thermal reservoir. $\mathcal{W}$, $\mathcal{S}$, $\mathcal{D}$ and $\mathcal{R}$ are initially prepared in states $\rho_{\mathcal{W}}$, $\rho_{\mathcal{S}}$, $|\psi\rangle$, and $\tau^{\beta}_{\mathcal{R}}$ respectively. (I) Measurement: first, $\mathcal{S}$ and $\mathcal{D}$ are coupled by the joint premeasurement unitary U_M. This correlates the two systems so that the measurement outcomes of the observable $M_{\mathcal{S}}$ on $\mathcal{S}$, namely $x \in \mathcal{X}$, are 'stored' in $\mathcal{D}$'s memory as the eigenstates of an observable $Z_{\mathcal{D}}$. These are the states $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$. The system's post-measurement state $|\tilde{\varphi}_x\rangle$ will be classically correlated with the demon memory state $|\psi_x\rangle$, occurring with a probability $p^M_{\rho_{\mathcal{S}}}(x)$. (II) Feedback: the global feedback unitary operator V then couples $\mathcal{W}$ and $\mathcal{S}$ such that, conditional on the outcome x, they evolve by the CPTP maps Λ^*_x and Λ_x, respectively. (III) Erasure: at the end of feedback, the demon's memory is erased by coupling to the thermal reservoir $\mathcal{R}$ with the unitary interaction U_R.

2. Modeling a quantum Szilard engine

A general quantum Szilard engine is constituted of four subsystems: a system $\mathcal{S}$; a demon $\mathcal{D}$; a weight $\mathcal{W}$; and a thermal reservoir $\mathcal{R}$. These have the Hilbert space $\mathcal{H} = \mathcal{H}_{\mathcal{W}} \otimes \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{D}} \otimes \mathcal{H}_{\mathcal{R}}$, and respectively the Hamiltonians $H_{\mathcal{W}}$, $H_{\mathcal{S}}$, $H_{\mathcal{D}}$, and $H_{\mathcal{R}}$. When describing operators that act non-trivially on only one subsystem, we shall omit identities on the other subsystems for simplicity. Furthermore, we shall only consider finite-dimensional Hilbert spaces. This model has in common with [45, 46] and [29, 38] that it includes respectively the weight and the demon's memory within the quantum description. As with the classical Szilard engine, each cycle of our quantum Szilard engine involves two stages, namely, measurement and feedback. Before $\mathcal{D}$ can perform measurements in the next cycle, its memory must first be erased. This is achieved by an appropriate interaction with $\mathcal{R}$. As the state of $\mathcal{S}$ can be different at the end of the cycle, then unlike the classical Szilard engine, the quantum Szilard engine is, strictly speaking, not cyclical. However, as will be shown, such non-cyclicality will not result in a violation of the second law.

All Szilard engines must satisfy the following two requirements. Here, we shall state them colloquially, but will offer mathematically precise formulations in the next two subsections.

Requirement 1. Both the measuring and feedback processes must be energy conserving on the total system.

This is necessary for all work sources to be explicitly accounted for; if either the measuring or feedback process does not conserve the energy of the total system, then it will require work from an outside source.

Requirement 2. If the demon's memory is in a state corresponding to a measurement outcome x, the feedback process must result in a closed evolution of the compound of system plus weight (and reservoir, if it is present). After feedback, the demon's memory must remain in the same state.

This is necessary in order to conform with the functioning of the classical Szilard engine described above. There, upon discovering the particle's location, the demon arranges the weight-and-pulley mechanism accordingly so as to facilitate work extraction. After making its arrangements, the weight, system, and reservoir evolve as a closed, mechanically isolated system, while the demon's memory is unaltered.

In the subsequent sections, we shall depart from the traditional set-up of the Szilard engine by altering the feedback stage; this will no longer involve $\mathcal{R}$, and the source of work will not be identified as heat from the reservoir, but rather the internal energy of the compound $\mathcal{S} + \mathcal{D}$. Each cycle of work extraction is depicted schematically in figure 2. Our work is similar in spirit to that of [33], except that we model both the weight and demon's memory as explicit quantum systems, and impose energy conservation on the measuring process.

2.1. Measurement stage

During the measurement stage, the demon $\mathcal{D}$ performs a measurement on $\mathcal{S}$, and by doing so prepares it in a state that is correlated with the measurement outcome. For now, we will restrict ourselves to standard, non-

degenerate projective measurements, and shall generalize to degenerate observables in appendix C.2. If $\mathcal{H}_S \simeq \mathbb{C}^d$, the observable can be represented as the self-adjoint operator

$$M_S = \sum_{x\in\mathcal{X}} x P_S[\varphi_x], \tag{2.1}$$

where $\mathcal{X} \coloneqq \{1, \ldots, d\}$ are the measurement outcomes. Here $P_S[\varphi_x] \equiv |\varphi_x\rangle\langle\varphi_x|$ is a projection on the vector $|\varphi_x\rangle \in \mathcal{H}_S$. We wish to model the measurement of M_S as resulting from a physical interaction between $\mathcal{S}$ and $\mathcal{D}$, so that the outcomes $\mathcal{X}$ are stored in the memory of $\mathcal{D}$ by the orthogonal set of states $\{|\psi_x\rangle \in \mathcal{H}_D\}_{x\in\mathcal{X}}$. Therefore, we describe the measurement model of M_S, as defined in equation (2.1), by the tuple $\mathscr{M} \coloneqq (\mathcal{H}_D, |\psi\rangle, U_M, Z_D)$ [47–51]. Here $|\psi\rangle \in \mathcal{H}_D$ is the initial state of $\mathcal{D}$; U_M is the *premeasurement* unitary interaction between $\mathcal{S}$ and $\mathcal{D}$, characterized by

$$U_M : |\varphi_x\rangle \otimes |\psi\rangle \mapsto |\tilde{\varphi}_x\rangle \otimes |\psi_x\rangle, \tag{2.2}$$

where $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ can be any set of vectors on $\mathcal{H}_S$, which do not have to be orthogonal; and

$$Z_D = \sum_{x\in\mathcal{X}} x P_D^x \tag{2.3}$$

is an observable on $\mathcal{D}$ with each outcome x corresponding to the same for M_S. Here, P_D^x is a projection operator of arbitrary rank, such that for all $x \in \mathcal{X}$, $|\psi_x\rangle \in P_D^x(\mathcal{H}_D)$. If $\mathcal{H}_D \simeq \mathcal{H}_S$, then $P_D^x = P_D[\psi_x]$.

For an arbitrary initial state ρ_S of $\mathcal{S}$, the total state of $\mathcal{S} + \mathcal{D}$ after premeasurement is

$$\rho_{S+D}^M \coloneqq U_M(\rho_S \otimes P_D[\psi])U_M^\dagger. \tag{2.4}$$

In order for the measuring process to leave a classical record of outcomes, the demon's memory must be *objectified* [52]. That is to say, after coupling $\mathcal{S}$ with $\mathcal{D}$ by the premeasurement unitary as defined by equation (2.2), thus preparing the entangled state ρ_{S+D}^M as defined in equation (2.4), we must prepare the statistical mixture

$$\begin{aligned}\rho_{S+D}^{M,O} &\coloneqq \sum_{x\in\mathcal{X}} P_D^x \rho_{S+D}^M P_D^x, \\ &= \sum_{x\in\mathcal{X}} p_{\rho_S}^M(x) P_S[\tilde{\varphi}_x] \otimes P_D[\psi_x],\end{aligned} \tag{2.5}$$

where

$$p_{\rho_S}^M(x) \coloneqq \mathrm{tr}[P_S[\varphi_x]\rho_S] \tag{2.6}$$

is the Born rule probability of observing outcome x, given a measurement of M_S on $\mathcal{S}$, prepared in state ρ_S. Equation (2.5) is a proper mixture, or a *Gemenge* $(p_{\rho_S}^M(x),\ P_S[\tilde{\varphi}_x] \otimes P_D[\psi_x])$, which can be interpreted as each state $P_S[\tilde{\varphi}_x] \otimes P_D[\psi_x]$ being prepared according to a probability distribution $p_{\rho_S}^M(x)$, as given by equation (2.6). Moreover, $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ can be interpreted as the set of post-measurement states on $\mathcal{S}$. We may objectify $\mathcal{D}$ by performing an unselective Lüders measurement of Z_D [50], as defined in equation (2.3), on $\mathcal{D}$. Alternatively, as shown in [38], $\mathcal{D}$ can be objectified by unitarily coupling it with an auxiliary system. In the subsequent section we show that imposing requirement 2 on the feedback process implies that it does not matter whether we objectify the demon before or after the feedback stage.

Definition 1. Consider a system with Hilbert space $\mathcal{H}$ and Hamiltonian H. The completely positive, trace preserving (CPTP) map $\mathcal{E}$ is said to conserve energy if

$$\mathrm{tr}[H\rho] = \mathrm{tr}[H\mathcal{E}(\rho)] \tag{2.7}$$

for all states ρ on $\mathcal{H}$.

Lemma 1. *The measuring process satisfies requirement 1, i.e., is energy conserving, if both* $[Z_D, H_D]_- = \mathbb{O}$ *and* $[U_M, H_S + H_D]_- = \mathbb{O}$, *where* H_S *and* H_D *are the system and demon Hamiltonians, respectively, and* Z_D *is the demon observable defined in equation (2.3).*

Proof. The measuring process consists of premeasurement and objectification. Given definition 1, these are energy conserving if

$$\mathrm{tr}[(H_S + H_D)\rho_{S+D}^{M,O}] = \mathrm{tr}[(H_S + H_D)\rho_S \otimes P_D[\psi]] \tag{2.8}$$

for all ρ_S on $\mathcal{H}_S$, where $\rho_{S+D}^{M,O}$ is given by equation (2.5). Therefore, we must have $[U_M, H_S + H_D]_- = \mathbb{O}$ and $[P_D^x, H_D]_- = \mathbb{O}$ for all $x \in \mathcal{X}$. The latter condition is equivalent to $[Z_D, H_D]_- = \mathbb{O}$. □

Now we may analyze feature 1 with respect to requirement 1.

Lemma 2. *Let the measuring process satisfy requirement 1. It follows that the measurement of $M_{\mathcal{S}}$, as defined by equation (2.1), will satisfy feature 1, i.e., it will be repeatable, if and only if the post-measurement states $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ are eigenvectors of $H_{\mathcal{S}}$.*

Proof. The post-measurement state of $\mathcal{S}$, conditional on outcome x, is $|\tilde{\varphi}_x\rangle$. The probability of observing outcome x in a subsequent measurement of $M_{\mathcal{S}}$ will be $p^{M}_{\tilde{\varphi}_x}(x) = |\langle\tilde{\varphi}_x|\varphi_x\rangle|^2$, as determined by equation (2.6). This equals unity if and only if $|\tilde{\varphi}_x\rangle = e^{i\theta}|\varphi_x\rangle$. Therefore, $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ must be eigenvectors of $M_{\mathcal{S}}$.

To show that $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ must be eigenvectors of $H_{\mathcal{S}}$ if the measurement is repeatable, we use the WAY theorem. The WAY theorem can be stated thusly: let the premeasurement unitary operator in the measurement model of $M_{\mathcal{S}}$, i.e., U_M, commute with $H_{\mathcal{S}} + H_{\mathcal{D}}$. If the measurement of $M_{\mathcal{S}}$ is repeatable, or $[Z_{\mathcal{D}}, H_{\mathcal{D}}]_- = \mathbb{O}$, where $Z_{\mathcal{D}}$ is defined in equation (2.3), then $[M_{\mathcal{S}}, H_{\mathcal{S}}]_- = \mathbb{O}$. We refer to [42] for a proof. If $M_{\mathcal{S}}$ commutes with $H_{\mathcal{S}}$, then they will share the same eigenvectors. □

2.2. Feedback stage

During the feedback stage, the demon brings the system in contact with the weight, $\mathcal{W}$, which is initially prepared in state $\rho_{\mathcal{W}}$. Conforming with requirement 2, the demon then evolves the compound system of $\mathcal{W} + \mathcal{S}$ by the unitary operator U_x, which is chosen conditional on the measurement outcome $x \in \mathcal{X}$. We wish to determine the global feedback unitary operator V that achieves this.

Lemma 3. *Feedback is implemented by a unitary operator V acting on the compound system $\mathcal{W} + \mathcal{S} + \mathcal{D}$. V will satisfy requirement 2 if and only if it can be written as*

$$V = \sum_{x\in\mathcal{X}} U_x \otimes P^x_{\mathcal{D}}, \tag{2.9}$$

such that U_x are unitary operators on $\mathcal{H}_{\mathcal{W}} \otimes \mathcal{H}_{\mathcal{S}}$, and $P^x_{\mathcal{D}}$ are the projection operators defined in equation (2.3).

Proof. Requirement 2 states that if the demon is in a state corresponding to a measurement outcome x, the system and weight must undergo a closed evolution. Consequently, V must satisfy

$$V(|\Psi\rangle \otimes |\psi_x\rangle) = (U_x|\Psi\rangle) \otimes |\psi_x\rangle \tag{2.10}$$

for all $x \in \mathcal{X}$ and $|\Psi\rangle \in \mathcal{H}_{\mathcal{W}} \otimes \mathcal{H}_{\mathcal{S}}$, where $|\psi_x\rangle$ is an eigenstate of the demon observable $Z_{\mathcal{D}}$ as defined in equation (2.3). This is clearly satisfied if V is of the form equation (2.9). To prove only if, we note that equation (2.10) implies that

$$V(|\Psi\rangle \otimes |\psi_x\rangle) = (P^x_{\mathcal{D}} V P^x_{\mathcal{D}})(|\Psi\rangle \otimes |\psi_x\rangle) \tag{2.11}$$

for all $x \in \mathcal{X}$, where $P^x_{\mathcal{D}}$ is a projection on the subspace of $\mathcal{H}_{\mathcal{D}}$ that contains $|\psi_x\rangle$. Therefore, it follows that

$$V = \sum_{x\in\mathcal{X}} P^x_{\mathcal{D}} V P^x_{\mathcal{D}}, \tag{2.12}$$

and so V must be of the form equation (2.9). □

Corollary 1. *Let the feedback unitary satisfy requirement 2. Then the state of the compound $\mathcal{W} + \mathcal{S} + \mathcal{D}$ will be identical whether $\mathcal{D}$ is objectified prior to feedback, or after it.*

Proof. The compound of $\mathcal{S} + \mathcal{D}$ after premeasurement and objectification is given by equation (2.5). After feedback, the state of the compound $\mathcal{W} + \mathcal{S} + \mathcal{D}$ is

$$V(\rho_{\mathcal{W}} \otimes \rho^{M,O}_{\mathcal{S}+\mathcal{D}})V^\dagger = V\left(\sum_{x\in\mathcal{X}} P^x_{\mathcal{D}}(\rho_{\mathcal{W}} \otimes \rho^{M}_{\mathcal{S}+\mathcal{D}})P^x_{\mathcal{D}}\right)V^\dagger. \tag{2.13}$$

If the feedback unitary is of the form equation (2.9), then $[V, P^x_{\mathcal{D}}]_- = \mathbb{O}$ for all $x \in \mathcal{X}$, and so we have

$$V\left(\sum_{x\in\mathcal{X}} P^x_{\mathcal{D}}(\rho_{\mathcal{W}} \otimes \rho^{M}_{\mathcal{S}+\mathcal{D}})P^x_{\mathcal{D}}\right)V^\dagger = \sum_{x\in\mathcal{X}} P^x_{\mathcal{D}} V(\rho_{\mathcal{W}} \otimes \rho^{M}_{\mathcal{S}+\mathcal{D}})V^\dagger P^x_{\mathcal{D}}. \tag{2.14}$$

The second line corresponds to performing feedback after premeasurement, but before objectification has occurred. □

We now show that if V as defined by equation (2.9) is to satisfy requirement 1, then each U_x must conserve $H_{\mathcal{W}} + H_{\mathcal{S}}$.

Lemma 4. *Let V be a feedback unitary operator that satisfies requirement 2. It will also satisfy requirement 1 if and only if: (i) $[U_x, H_{\mathcal{W}} + H_{\mathcal{S}}]_- = \mathbb{O}$ for all $x \in \mathcal{X}$; and (ii) for every subset $\mathcal{X}' \subseteq \mathcal{X}$ such that $U_x = U_y$ for all $x, y \in \mathcal{X}'$, $\sum_{x\in\mathcal{X}'}[P^x_{\mathcal{D}}, H_{\mathcal{D}}]_- = \mathbb{O}$.*

Proof. In order for V as defined by equation (2.9) to conserve the total energy, by definition 1 we require that

$$\mathrm{tr}[HV\rho V^\dagger] = \mathrm{tr}[H\rho] \tag{2.15}$$

for all states ρ on $\mathcal{H}_{\mathcal{W}} \otimes \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{D}}$, where $H = H_{\mathcal{W}} + H_{\mathcal{S}} + H_{\mathcal{D}}$. Therefore, V must commute with the total Hamiltonian. Because of the additivity of the Hamiltonian, $[V, H]_- = \mathbb{O}$ can be written as

$$\sum_{x\in\mathcal{X}} [U_x, H_{\mathcal{W}} + H_{\mathcal{S}}]_- \otimes P^x_{\mathcal{D}} = -\sum_{x\in\mathcal{X}} U_x \otimes [P^x_{\mathcal{D}}, H_{\mathcal{D}}]_-. \tag{2.16}$$

Given an arbitrary pair of states $|\psi_x\rangle \in P^x_{\mathcal{D}}(\mathcal{H}_{\mathcal{D}})$ and $|\psi_y\rangle \in P^y_{\mathcal{D}}(\mathcal{H}_{\mathcal{D}})$, such that $x \neq y$, and referring to the right hand and left hand sides of equation (2.16) as RHS and LHS, respectively, we see that

$$\begin{aligned} \langle\psi_x|\mathrm{LHS}|\psi_y\rangle &= \mathbb{O}, \\ \langle\psi_x|\mathrm{RHS}|\psi_y\rangle &= \langle\psi_x|H_{\mathcal{D}}|\psi_y\rangle\,(U_y - U_x). \end{aligned} \tag{2.17}$$

However, given equation (2.16), we must have $\langle\psi_x|\mathrm{LHS}|\psi_y\rangle = \langle\psi_x|\mathrm{RHS}|\psi_y\rangle$. This is satisfied if either: (i) $[P^z_{\mathcal{D}}, H_{\mathcal{D}}]_- = \mathbb{O}$ for $z \in \{x, y\}$; or (ii) $U_x = U_y$. Option (i) satisfies the if statement of the lemma. Option (ii) implies that equation (2.16) is satisfied if

$$[U_{\mathcal{X}'}, H_{\mathcal{W}} + H_{\mathcal{S}}]_- \otimes P^{\mathcal{X}'}_{\mathcal{D}} = -U_{\mathcal{X}'} \otimes [P^{\mathcal{X}'}_{\mathcal{D}}, H_{\mathcal{D}}]_- \tag{2.18}$$

for all maximal subsets $\mathcal{X}' \subseteq \mathcal{X}$ such that, given all $x, y \in \mathcal{X}'$, $U_x = U_y = U_{\mathcal{X}'}$. Here we define $P^{\mathcal{X}'}_{\mathcal{D}} := \sum_{x\in\mathcal{X}'} P^x_{\mathcal{D}}$.

Equation (2.18) is satisfied if: (a) $[P^{\mathcal{X}'}_{\mathcal{D}}, H_{\mathcal{D}}]_- \propto P^{\mathcal{X}'}_{\mathcal{D}}$ and $[U_{\mathcal{X}'}, H_{\mathcal{W}} + H_{\mathcal{S}}]_- \propto U_{\mathcal{X}'}$; or (b) if $[P^{\mathcal{X}'}_{\mathcal{D}}, H_{\mathcal{D}}]_- = \mathbb{O}$ and $[U_{\mathcal{X}'}, H_{\mathcal{W}} + H_{\mathcal{S}}]_- = \mathbb{O}$. It is easy to verify that (a) is impossible, and so only option (b) is available. This concludes the proof of the only if portion of the lemma. □

For each measurement outcome x, as a result of the global feedback unitary operator V given in equation (2.9), $\mathcal{S}$ and $\mathcal{W}$ undergo the complementary CPTP maps

$$\begin{aligned} \Lambda_x : P_{\mathcal{S}}[\tilde{\varphi}_x] &\mapsto \mathrm{tr}_{\mathcal{W}}[U_x(\rho_{\mathcal{W}} \otimes P_{\mathcal{S}}[\tilde{\varphi}_x])\,U_x^\dagger], \\ \Lambda_x^* : \rho_{\mathcal{W}} &\mapsto \mathrm{tr}_{\mathcal{S}}[U_x(\rho_{\mathcal{W}} \otimes P_{\mathcal{S}}[\tilde{\varphi}_x])\,U_x^\dagger], \end{aligned} \tag{2.19}$$

where we recall that $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ are the post-measurement states of $\mathcal{S}$.

We now wish to define the (average) work that is transferred from $\mathcal{S}$ into $\mathcal{W}$, for each measurement outcome, as a result of feedback. To this end, we use the following definition.

Definition 2. For each measurement outcome $x \in \mathcal{X}$, the average work transferred into the weight is defined as

$$W_x := F(\Lambda_x^*[\rho_{\mathcal{W}}]) - F(\rho_{\mathcal{W}}), \tag{2.20}$$

where: Λ_x^* is the CPTP map defined by equation (2.19);

$$F(\rho) := \mathrm{tr}[H\rho] - K_{\mathrm{B}}T\,S(\rho) \tag{2.21}$$

is the non-equilibrium free energy of a system with state ρ, relative to the Hamiltonian H and temperature T; and $S(\rho) := -\mathrm{tr}[\rho\ln(\rho)]$ is the von-Neumann entropy of ρ.

This definition has been argued for previously in [53, 54]. Even though the thermal reservoir is not involved during feedback, it is still part of the thermodynamic context of the Szilard engine. As such, work can be extracted from both the system, and the weight, by letting them interact appropriately with the reservoir. Therefore, the quantifier of work transfer must be temperature dependent, in the form of free energy difference, in order to: (i) ensure consistency with the 'internal' description of work extraction from $\mathcal{S}$, wherein the weight is not included in the quantum description; and (ii) avoid violation of the second law. For a detailed argument we refer the reader to appendix A. We note that an alternative definition for work transfer to the weight is the increase in the internal energy of $\mathcal{W}$. While this formulation will be consistent with the second law only if the feedback unitary V induces unital dynamics on the system $\mathcal{S}$ [55], definition 2 does not suffer from such limitations. Moreover, definition 2 reduces to the increase in internal energy when feature 2 is satisfied.

Now that we have defined work extraction, we may analyze this with respect to feature 2.

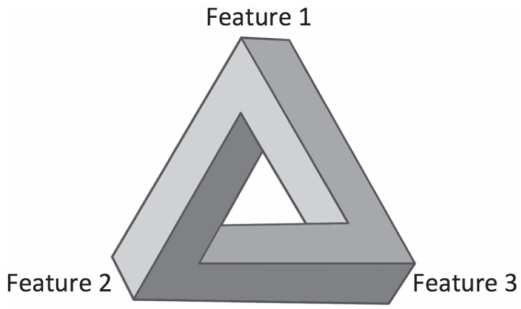

Figure 3. The impossible triangle of a quantum Szilard engine powered by projective measurements. Features 1, 2, and 3 signify respectively the repeatability of the measurement; invariant weight entropy; and the reliability of the engine. The fact that only two vertices of the impossible triangle can be physically connected, but not the third, represents the result that all three features cannot be simultaneously satisfied.

Definition 3. The Szilard engine satisfies feature 2 if for all $x \in \mathcal{X}$,

$$S(\Lambda_x^*[\rho_{\mathcal{W}}]) = S(\rho_{\mathcal{W}}). \tag{2.22}$$

Lemma 5. *When the Szilard engine satisfies feature 2, it follows that*

$$W_x \leqslant \langle\tilde{\varphi}_x|H_{\mathcal{S}}|\tilde{\varphi}_x\rangle - \min[\sigma(H_{\mathcal{S}})], \tag{2.23}$$

where $\sigma(H_{\mathcal{S}})$ is the spectrum of $H_{\mathcal{S}}$.

Proof. The work transferred into $\mathcal{W}$ is, by definition 2 and lemma 4, given as

$$\begin{aligned} W_x \coloneqq \mathrm{tr}[H_{\mathcal{S}}(P_{\mathcal{S}}[\tilde{\varphi}_x] - \Lambda_x[\tilde{\varphi}_x])] \\ + K_{\mathrm{B}}T(S(\rho_{\mathcal{W}}) - S(\Lambda_x^*[\rho_{\mathcal{W}}])). \end{aligned} \tag{2.24}$$

As $\mathrm{tr}[H_{\mathcal{S}}\Lambda_x[\tilde{\varphi}_x]] \geqslant \min[\sigma(H_{\mathcal{S}})]$, it follows that

$$\begin{aligned} W_x \leqslant \langle\tilde{\varphi}_x|H_{\mathcal{S}}|\tilde{\varphi}_x\rangle - \min[\sigma(H_{\mathcal{S}})] \\ + K_{\mathrm{B}}T(S(\rho_{\mathcal{W}}) - S(\Lambda_x^*[\rho_{\mathcal{W}}])). \end{aligned} \tag{2.25}$$

If the Szilard engine satisfies feature 2, then by definition 3 we have equation (2.23). □

3. The impossibility theorem

We are now ready to prove a main result of this paper. The impossibility theorem is illustrated by Penrose's impossible triangle in figure 3.

Theorem 1. *Consider a quantum Szilard engine that, during the feedback stage, operates in thermal isolation. Let the engine satisfy requirements 1 and 2. It follows that if the engine satisfies any two from features 1, 2, and 3, it will necessarily fail to satisfy the third.*

Proof. Let the engine satisfy features 1 and 2. By lemma 2 the post-measurement states $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ are the eigenvectors of $M_{\mathcal{S}}$ and, hence, $H_{\mathcal{S}}$. Consequently, for some outcome $x \in \mathcal{X}$, $\langle\tilde{\varphi}_x|H_{\mathcal{S}}|\tilde{\varphi}_x\rangle = \min[\sigma(H_{\mathcal{S}})]$. By lemma 5, for this outcome we have $W_x \leqslant 0$, and feature 3 cannot be satisfied.

Let the engine satisfy features 1 and 3. By lemma 2 the post-measurement states $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ are the eigenvectors of $M_{\mathcal{S}}$ and, hence, $H_{\mathcal{S}}$. Consequently, for some outcome $x \in \mathcal{X}$, $\langle\tilde{\varphi}_x|H_{\mathcal{S}}|\tilde{\varphi}_x\rangle = \min[\sigma(H_{\mathcal{S}})]$. By lemma 5, for this outcome $W_x > 0$ only if $S(\Lambda_x^*[\rho_{\mathcal{W}}]) < S(\rho_{\mathcal{W}})$. Hence, feature 2 cannot be satisfied.

Let the engine satisfy features 2 and 3. By lemma 5, for all $x \in \mathcal{X}$, the work is bounded as $W_x \leqslant \langle\tilde{\varphi}_x|H_{\mathcal{S}}|\tilde{\varphi}_x\rangle - \min[\sigma(H_{\mathcal{S}})]$. As $W_x > 0$ for all $x \in \mathcal{X}$, it follows that $\langle\tilde{\varphi}_x|H_{\mathcal{S}}|\tilde{\varphi}_x\rangle > \min[\sigma(H_{\mathcal{S}})]$ for all $x \in \mathcal{X}$. Therefore, the post-measurement states $\{|\tilde{\varphi}_x\rangle\}_{x\in\mathcal{X}}$ cannot be the eigenvectors of $H_{\mathcal{S}}$. By lemma 2, feature 1 cannot be satisfied. □

Theorem 1, simply stated, says that if the system is measured with respect to a non-degenerate observable, in a repeatable and energy conserving fashion, it must be projected onto the eigenstates of $H_{\mathcal{S}}$. Consequently, if we do not allow the weight's entropy to decrease, then for the outcome that projects the system onto the groundstate of $H_{\mathcal{S}}$, zero work can be extracted.

In appendix B, we illustrate the incompatibility between the three features by looking at a concrete model where both $\mathcal{S}$ and $\mathcal{D}$ are qubits, while $\mathcal{W}$ is a harmonic oscillator. In appendix C we show that theorem 1 can be circumvented if: (i) the thermal reservoir is involved during the feedback stage so that, just as in the classical Szilard engine, the source of work will be heat drawn from the reservoir; or (ii) the observable measured on $\mathcal{S}$ is degenerate and is measured 'inefficiently'.

4. Net work extraction per cycle

Figure 2 depicts a single cycle of the Szilard engine under consideration. In appendix D we evaluate the net work extraction per cycle, wherein we do not distinguish between measurement outcomes. Labeling the 'coarse-grained' work transferred to the weight as $W_{\mathcal{X}} \coloneqq F(\rho'_{\mathcal{W}}) - F(\rho_{\mathcal{W}})$, and the work cost of erasure as W_R, the net coarse-grained work is shown to obey the inequality

$$W_{\mathcal{X}}^{\text{net}} \coloneqq W_{\mathcal{X}} - W_R \leqslant F(\rho_{\mathcal{S}}) - F(\rho'_{\mathcal{S}}), \tag{4.1}$$

where $\rho'_{\mathcal{S}}$ and $\rho'_{\mathcal{W}}$ are the average states of $\mathcal{S}$ and $\mathcal{W}$ at the end of the cycle, respectively, obtained by sampling the states $\Lambda_x(\tilde{\varphi}_x)$ and $\Lambda_x^*(\rho_{\mathcal{W}})$ by the probability distribution $p^M_{\rho_{\mathcal{S}}}(x)$ as defined by equation (2.6). We note that equation (4.1) holds irrespective of whether the Szilard engine satisfies any of features 1, 2, or 3. Moreover, we note that the coarse-grained work is generally smaller than the average work, i.e., $W_{\mathcal{X}} \leqslant \langle W_x \rangle \coloneqq \sum_{x\in\mathcal{X}} p^M_{\rho_{\mathcal{S}}}(x)\, W_x$, where W_x is defined in equation (2.20). While the coarse-grained work extraction obeys the second law, the average work will not; if $\rho_{\mathcal{S}}$ is thermal, then $W_{\mathcal{X}}^{\text{net}} \leqslant 0$ whereas $\langle W_x^{\text{net}} \rangle \coloneqq \langle W_x \rangle - W_R$ can be positive.

To be sure, the second law is a statistical statement, held true precisely when we do not have access to the individual measurement outcomes. Let us recall the definition for work transferred into the weight when it transforms as $\rho_{\mathcal{W}} \mapsto \Lambda_x^*(\rho_{\mathcal{W}})$, given by definition 2 and articulated in appendix A. This was given operational meaning as being the maximum value of work that can be extracted from the weight, by an isothermal process $\Lambda_x^*(\rho_{\mathcal{W}}) \mapsto \rho_{\mathcal{W}}$ involving the reservoir of temperature T. However, if we were to forget the measurement outcomes, then we could not use such information to tailor our process of extracting work from the weight. Indeed, this protocol must be designed with only the average state of the weight in mind. The maximum value of work extractable from the weight, given an isothermal process $\rho'_{\mathcal{W}} \mapsto \rho_{\mathcal{W}}$, is precisely $W_{\mathcal{X}}$.

5. Discussion

We give a general mathematical description of a quantum Szilard engine that operates in two stages, namely, projective measurement and feedback. In our model, in contradistinction to the classical Szilard engine, the feedback stage does not involve the thermal reservoir. Here, the source of work is the energetic changes due to (non-degenerate) projective measurements. In order to avoid cheating by the inclusion of hidden work sources, we impose energy conservation on the measuring process. As a result of the WAY theorem, the observables that the demon can measure will be limited to those that commute with the system's Hamiltonian.

We showed that while the Szilard engine, in lieu of a thermal reservoir, can be powered by (non-degenerate) projective measurements, it cannot simultaneously satisfy three features of the classical Szilard engine model; the conjunction of any two will preclude the possibility of the third. These features are: (i) the measurement performed by the demon is repeatable, meaning that conditional on obtaining outcome x, a subsequent measurement of the same observable would yield x with certainty; (ii) the weight's entropy does not change as a result of feedback; and (iii) work extraction is reliable, i.e., is strictly positive for all measurement outcomes. This observation is a first step towards developing 'second-law-like' relations in the context of measurement-assisted feedback control beyond thermality. While the second law results from entropic considerations, these 'second-law-like' relations would result from energy conservation of unitary interactions that implement measurements.

The Szilard engine here discussed is, strictly speaking, not cyclical; at the end of a cycle of work extraction, the state of the system, $\rho'_{\mathcal{S}}$, will not be the same as its initial state, $\rho_{\mathcal{S}}$. For the engine to be made cyclical, therefore, we must have at our disposal an infinite supply of systems with state $\rho_{\mathcal{S}}$ such that, at the end of each cycle, the system's state is swapped with one of these. One example of such 'free resources' is if $\rho_{\mathcal{S}}$ is thermal. Here, we may interpret the closure of the cycle to result from the system being brought to thermal equilibrium with the reservoir.

The strict non-cyclicality of the engine notwithstanding, the statistical second law will not be violated. This is because, when taking the erasure cost of the demon into consideration, the total net work extracted from the system will be bounded by the decrease in its free energy—a quantity that will not be positive if the system is initially at thermal equilibrium. However, this requires a careful consideration of how one should evaluate work when choosing to 'forget' the measurement outcomes—precisely the domain where the second law is applicable. As with unselective measurements, the work transferred to the weight when the indvidual measurement

outcomes are not distinguished from one another must be defined by how the weight's state changes *on average*. Indeed, the extractable work from the weight, when the measurement outcomes are forgotten, is smaller than the average value of work, when the measurement outcomes are taken into consideration.

Acknowledgments

The authors would like to thank L D Loveridge, K Abdelkhalek, D Reeb, K Hovhannisyan, and H Miller for the useful discussions that helped in developing the ideas presented in this paper. JA acknowledges support from EPSRC, grant EP/M009165/1, and the Royal Society. This research was supported by the COST network MP1209 'Thermodynamics in the quantum regime'.

Appendix A. Definition of work transferred into the weight

Here we wish to justify defining the work transferred into the weight, as a result of feedback, by definition 2. To this end, let us first recall a known result from standard non-equilibrium quantum thermodynamics. In the *internal* description of work extraction, in contradistinction to the *external* description, the weight is not included in the quantum formalism. Here, the work extracted from a system undergoing a (non-energy conserving) unitary evolution is defined as the decrease in its internal energy. Consequently, if a system $\mathcal{S}$ undergoes a transformation $\rho_{\mathcal{S}} \mapsto \rho'_{\mathcal{S}} \coloneqq \mathrm{tr}_{\mathcal{R}}[U(\rho_{\mathcal{S}} \otimes \tau^{\beta}_{\mathcal{R}})U^{\dagger}]$, where U is a global unitary operator and $\tau^{\beta}_{\mathcal{R}} \coloneqq \mathrm{e}^{-\beta H_{\mathcal{R}}}/\mathrm{tr}[\mathrm{e}^{-\beta H_{\mathcal{R}}}]$ is the thermal state of the thermal reservoir $\mathcal{R}$, with $\beta \coloneqq (K_{\mathrm{B}}T)^{-1}$ the inverse temperature, the work extracted obeys the inequality

$$W_{\mathrm{ext}}(\rho_{\mathcal{S}} \mapsto \rho'_{\mathcal{S}}) \coloneqq \mathrm{tr}[(H_{\mathcal{S}} + H_{\mathcal{R}})\rho_{\mathcal{S}} \otimes \tau^{\beta}_{\mathcal{R}}] - \mathrm{tr}[(H_{\mathcal{S}} + H_{\mathcal{R}})U(\rho_{\mathcal{S}} \otimes \tau^{\beta}_{\mathcal{R}})U^{\dagger}] \leqslant F(\rho_{\mathcal{S}}) - F(\rho'_{\mathcal{S}}), \tag{A.1}$$

with the equality obtained when the interaction between system and thermal reservoir is 'quasi-static' [56].

Therefore, definition 2 can be justified with the following argument. When the weight interacts with the system, thereby transforming as $\rho_{\mathcal{W}} \mapsto \Lambda^{*}_{x}(\rho_{\mathcal{W}})$, where Λ^{*}_{x} is given by equation (2.19), work is *transferred* to it. We may then perform the reverse transformation on the weight, i.e., $\Lambda^{*}_{x}(\rho_{\mathcal{W}}) \mapsto \rho_{\mathcal{W}}$, by an appropriate unitary interaction with the thermal reservoir, so as to extract this work. The work extracted here will be in the internal description, as there is no second weight into which the work is being transferred. By equation (A.1), the work we may extract obeys the inequality

$$W_{\mathrm{ext}}(\Lambda^{*}_{x}(\rho_{\mathcal{W}}) \mapsto \rho_{\mathcal{W}}) \leqslant F(\Lambda^{*}_{x}[\rho_{\mathcal{W}}]) - F(\rho_{\mathcal{W}}). \tag{A.2}$$

Clearly, the work transferred into the weight must be at least as great as the work that can be extracted from the weight, i.e.,

$$W_{x} \geqslant W_{\mathrm{ext}}(\Lambda^{*}_{x}(\rho_{\mathcal{W}}) \mapsto \rho_{\mathcal{W}}). \tag{A.3}$$

A natural assumption to make is that, since the process of transferring work into the weight is independent of the process by which work is extracted from the weight, the right hand side of the above equation should be replaced by the upper bound of equation (A.2). If we also take the view that transferring more work into the weight than can possibly be extracted from it is physically meaningless, we arrive at definition 2.

We also note that definition 2 is consistent with the internal description of work from the system $\mathcal{S}$, and that it satisfies the second law.

Lemma 6. *Let the system and weight be initially prepared in the states $\rho_{\mathcal{S}}$ and $\rho_{\mathcal{W}}$, respectively. Let the two systems evolve by a unitary operator U that conserves the total Hamiltonian $H_{\mathcal{W}} + H_{\mathcal{S}}$, and induces the complementary CPTP maps Λ on $\mathcal{S}$ and Λ^{*} on $\mathcal{W}$. Then the work transferred into the weight, W, as defined by definition 2, will never exceed the maximum work that can be directly extracted from the system by the process $\rho_{\mathcal{S}} \mapsto \Lambda(\rho_{\mathcal{S}})$, in the internal description, and using a single thermal reservoir at temperature T. If $\rho_{\mathcal{S}}$ is thermal, then W cannot be positive.*

Proof. By definition 2, energy conservation of U, and the subadditivity of the von-Neumann entropy, we have

$$\begin{aligned} W &\coloneqq F(\Lambda^{*}[\rho_{\mathcal{W}}]) - F(\rho_{\mathcal{W}}), \\ &= \mathrm{tr}[H_{\mathcal{W}}(\Lambda^{*}[\rho_{\mathcal{W}}] - \rho_{\mathcal{W}})] + K_{\mathrm{B}}T\,(S(\varrho_{\mathcal{W}}) - S(\Lambda^{*}[\rho_{\mathcal{W}}])), \\ &= \mathrm{tr}[H_{\mathcal{S}}(\rho_{\mathcal{S}} - \Lambda[\rho_{\mathcal{S}}])] \\ &\quad + K_{\mathrm{B}}T\,(S(\varrho_{\mathcal{W}}) - S(\Lambda^{*}[\rho_{\mathcal{W}}])), \\ &\leqslant \mathrm{tr}[H_{\mathcal{S}}(\rho_{\mathcal{S}} - \Lambda[\rho_{\mathcal{S}}])] + K_{\mathrm{B}}T\,(S(\Lambda[\rho_{\mathcal{S}}]) - S(\rho_{\mathcal{S}})), \\ &= F(\rho_{\mathcal{S}}) - F(\Lambda[\rho_{\mathcal{S}}]). \end{aligned} \tag{A.4}$$

By equation (A.1), we see that W is never greater than the upper bound of $W_{\text{ext}}(\rho_{\mathcal{S}} \mapsto \Lambda[\rho_{\mathcal{S}}])$. Moreover, if the system is initially in the thermal state $\rho_{\mathcal{S}} = \rho_{\mathcal{S}}^{\beta} := \mathrm{e}^{-\beta H_{\mathcal{S}}}/\mathrm{tr}[\mathrm{e}^{-\beta H_{\mathcal{S}}}]$, we have

$$\begin{aligned} W &\leqslant F(\rho_{\mathcal{S}}^{\beta}) - F(\Lambda[\rho_{\mathcal{S}}^{\beta}]), \\ &= -K_B T\, S(\Lambda[\rho_{\mathcal{S}}^{\beta}] \| \rho_{\mathcal{S}}^{\beta}), \end{aligned} \tag{A.5}$$

where $S(\rho\|\sigma) := \mathrm{tr}[\rho(\ln(\rho) - \ln(\sigma))]$ is the entropy of ρ relative to σ, which is a non-negative number and vanishes if and only if $\rho = \sigma$. Therefore, $W \leqslant 0$. □

Appendix B. An example with qubits

As an illustrative example, consider the simple case where $\mathcal{S}$ and $\mathcal{D}$ are both qubits, with the Hamiltonians

$$\begin{aligned} H_{\mathcal{S}} &:= \frac{\omega}{2}(P_{\mathcal{S}}[\varphi_+] - P_{\mathcal{S}}[\varphi_-]), \\ H_{\mathcal{D}} &:= \lambda_+ P_{\mathcal{D}}[\psi_+] + \lambda_- P_{\mathcal{D}}[\psi_-]. \end{aligned} \tag{B.1}$$

Furthermore, let the initial state of the system be

$$\rho_{\mathcal{S}} = q P_{\mathcal{S}}[\varphi_+] + (1 - q) P_{\mathcal{S}}[\varphi_-], \tag{B.2}$$

while that of $\mathcal{D}$ is $|\psi\rangle$. We wish to measure a two-valued observable $M_{\mathcal{S}}$, with outcomes $\pm$, with the measurement model $\mathscr{M} = (\mathcal{H}_{\mathcal{D}}, |\psi\rangle, U_M, Z_{\mathcal{D}})$. In order to satisfy requirement 1 for the measuring process, as shown by lemmas 1 and 2, $M_{\mathcal{S}}$ and $Z_{\mathcal{D}}$ must commute with $H_{\mathcal{S}}$ and $H_{\mathcal{D}}$, respectively. Therefore, we choose

$$M_{\mathcal{S}} := \sum_{x \in \pm} x P_{\mathcal{S}}[\varphi_x], \tag{B.3}$$

and

$$Z_{\mathcal{D}} = \sum_{x \in \pm} x P_{\mathcal{D}}[\psi_x]. \tag{B.4}$$

Given our choice of $M_{\mathcal{S}}$ and $Z_{\mathcal{D}}$, the premeasurement unitary operator is chosen as

$$U_M : |\varphi_\pm\rangle \otimes |\psi\rangle \mapsto |\tilde{\varphi}_\pm\rangle \otimes |\psi_\pm\rangle. \tag{B.5}$$

Finally, in order for the engine to satisfy requirements 1 and 2 for the feedback process, we choose the global feedback unitary operator

$$V = \sum_{x \in \pm} U_x \otimes P_{\mathcal{D}}[\psi_\pm]. \tag{B.6}$$

Following [46], we will use a harmonic oscillator of frequency ω as the weight, with the Hamiltonian

$$H_{\mathcal{W}} := \omega \sum_{n \in \mathbb{N}} n\, P_{\mathcal{W}}[n]. \tag{B.7}$$

Consequently, the conditional work extraction unitaries on $\mathcal{W} + \mathcal{S}$, namely, $U_\pm$, can be constructed as

$$\begin{aligned} U_\pm := \sum_{n=3}^{\infty} \sum_{a,b \in \{\pm\}} &|n - f\rangle\langle n - g| \otimes |\varphi_a\rangle\langle\varphi_b|\langle\varphi_a|G_\pm|\varphi_b\rangle \\ &+ P_{\mathcal{W}}[1] \otimes \mathbb{1}_{\mathcal{S}}, \end{aligned} \tag{B.8}$$

where $f := \max\{1, a1 - b1\}$ and $g := \max\{1, b1 - a1\}$, with $a, b \in \{\pm\}$. Here, $G_\pm := |\varphi_-\rangle\langle\tilde{\varphi}_\pm| + |\varphi_+\rangle\langle\tilde{\varphi}_\pm^{\perp}|$ is a unitary operator on $\mathcal{S}$, such that $\langle\tilde{\varphi}_\pm|\tilde{\varphi}_\pm^{\perp}\rangle = 0$. Therefore, when the system undergoes a transition $|\varphi_+\rangle \mapsto |\varphi_-\rangle$, the weight eigenstates are shifted up by one quantum, and vice versa.

It can be easily verified that $[U_\pm, H_{\mathcal{W}} + H_{\mathcal{S}}]_- = \mathbb{O}$, even when $|\tilde{\varphi}_\pm\rangle$ are not eigenstates of the system Hamiltonian. If the weight is initialized in a pure state $\rho_{\mathcal{W}} := P_{\mathcal{W}}[\Psi]$, where $|\Psi\rangle$ is an equal superposition of N Hamiltonian eigenstates,

$$|\Psi\rangle := \frac{1}{\sqrt{N}} \sum_{n=2}^{N+1} |n\rangle, \tag{B.9}$$

then it can function as a work storage device. This is a result of the energy-translational invariance of $|\Psi\rangle$; adding or removing one quantum is identical to a coordinate transformation $n \mapsto n + 1$ and $n \mapsto n - 1$, respectively. Moreover, if $|\tilde{\varphi}_\pm\rangle$ are the eigenvectors of $H_{\mathcal{S}}$, then irrespective of N the resulting dynamics on both $\mathcal{S}$ and $\mathcal{W}$ will be unitary. As such, feature 2 will be satisfied in this case. This is not so when $|\tilde{\varphi}_\pm\rangle$ are superpositions of $H_{\mathcal{S}}$ eigenvectors. For example, in the case of $|\tilde{\varphi}_\pm\rangle = \frac{1}{\sqrt{2}}(|\varphi_+\rangle \pm |\varphi_-\rangle)$, we have

$$\langle\varphi_-|\Lambda_\pm(\tilde{\varphi}_\pm)|\varphi_-\rangle = \frac{2N-1}{2N}, \tag{B.10}$$

with

$$S(\Lambda^*_\pm(\rho_\mathcal{W})) < \frac{1}{2N}\ln(2N) + \frac{2N-1}{2N}\ln\left(\frac{2N}{2N-1}\right). \tag{B.11}$$

In the limit as N tends to infinity, the increase in the weight's entropy can be made arbitrarily small, thus approximately satisfying feature 2.

We now look at two possible implementations of measurement-assisted work extraction, labeled I and II. In I, the observable $M_\mathcal{S}$ is measured repeatably, thus satisfying feature 1, while in II this is not the case. As the weight is initially pure, its entropy can never decrease. Therefore, feature 3 is satisfied in II, but not in I.

B.1. Example I: repeatable measurement

Let $|\tilde{\varphi}_\pm\rangle = |\varphi_\pm\rangle$, thus satisfying feature 1. Consequently, the state of $\mathcal{S}+\mathcal{D}$ after premeasurement is

$$U_M(\rho_\mathcal{S}\otimes P_\mathcal{D}[\psi])U_M^\dagger = qP_{\mathcal{W}+\mathcal{S}}[\varphi_+\otimes\psi_+] + (1-q)P_{\mathcal{W}+\mathcal{S}}[\varphi_-\otimes\psi_-]. \tag{B.12}$$

Transforming this state with the weight by the global unitary V prepares

$$\mathrm{tr}_\mathcal{W}[VU_M(P_\mathcal{W}[\Psi]\otimes\rho_\mathcal{S}\otimes P_\mathcal{D}[\psi])U_M^\dagger V^\dagger] = qP_{\mathcal{W}+\mathcal{S}}[\varphi_-\otimes\psi_+] + (1-q)P_{\mathcal{W}+\mathcal{S}}[\varphi_-\otimes\psi_-]. \tag{B.13}$$

Comparing equation (B.12) with (B.13), we see that, as a result of feedback, the system undergoes the transition $|\varphi_+\rangle \mapsto |\varphi_-\rangle$ when the demon is in the state $|\psi_+\rangle$, resulting in a work extraction of ω. When the demon is in the state $|\psi_-\rangle$, on the other hand, the system was already in the groundstate $|\varphi_-\rangle$ and is left the same, resulting in zero work extraction. Therefore, feature 3 is not satisfied.

B.2. Example II: non-repeatable measurement

Let $|\tilde{\varphi}_\pm\rangle = \frac{1}{\sqrt{2}}(|\varphi_+\rangle \pm |\varphi_-\rangle)$. Hence, feature 1 is not satisfied. Consequently, the state of $\mathcal{S}+\mathcal{D}$ after premeasurement is

$$U_M(\rho_\mathcal{S}\otimes P_\mathcal{D}[\psi])U_M^\dagger = qP_{\mathcal{W}+\mathcal{S}}\left[\frac{(\varphi_+ + \varphi_-)}{\sqrt{2}}\otimes\psi_+\right] + (1-q)P_{\mathcal{W}+\mathcal{S}}\left[\frac{(\varphi_+ - \varphi_-)}{\sqrt{2}}\otimes\psi_-\right]. \tag{B.14}$$

Transforming this state with the weight by the global unitary V prepares, in the ideal limit of $N\to\infty$,

$$\mathrm{tr}_\mathcal{W}[VU_M(P_\mathcal{W}[\Psi]\otimes\rho_\mathcal{S}\otimes P_\mathcal{D}[\psi])U_M^\dagger V^\dagger] = qP_{\mathcal{W}+\mathcal{S}}[\varphi_-\otimes\psi_+] + (1-q)P_{\mathcal{W}+\mathcal{S}}[\varphi_-\otimes\psi_-]. \tag{B.15}$$

Comparing equation (B.14) with (B.15) we see that, as a result of feedback, the system undergoes the transition $\frac{1}{\sqrt{2}}(|\varphi_+\rangle \pm |\varphi_-\rangle) \mapsto |\varphi_-\rangle$ when the demon is in the states $|\psi_\pm\rangle$, resulting in a work extraction of $\omega/2$ for both measurement outcomes. Therefore, feature 3 is satisfied.

Appendix C. Satisfying all three features with either a thermal reservoir, or degenerate observables

There are at least two ways in which theorem 1 can be circumvented: (i) letting the reservoir $\mathcal{R}$ be involved during the feedback stage; and (ii) measure $\mathcal{S}$ with a degenerate observable.

C.1. Szilard engine with heat from a thermal reservoir

As a simple example, let $\mathcal{S}$ be a d-dimensional system, and let $\mathcal{R}$ be a system initially prepared in the thermal state

$$\tau^\beta_\mathcal{R} := \frac{\mathrm{e}^{-\beta H_\mathcal{R}}}{\mathrm{tr}[\mathrm{e}^{-\beta H_\mathcal{R}}]}, \tag{C.1}$$

where $\beta = (K_\mathrm{B}T)^{-1}$ is the inverse temperature. By lemma 2, the non-degenerate observable $M_\mathcal{S} = \sum_{x\in\mathcal{X}} P[\varphi_x]$ can only be measured repeatably if it commutes with the system Hamiltonian $H_\mathcal{S}$. As such, in order to satisfy feature 1 the post-measurement states $\{|\varphi_x\rangle\}_{x\in\mathcal{X}}$ must be eigenstates of $H_\mathcal{S}$. Including the reservoir in the feedback stage means that the U_x in the feedback unitary operator defined in equation (2.9) are unitary operators on the compound $\mathcal{W}+\mathcal{S}+\mathcal{R}$ such that $[U_x, H_\mathcal{W}+H_\mathcal{S}+H_\mathcal{R}]_- = \mathbb{O}$. The CPTP maps defined in equation (2.19) will therefore be modified as

$$\begin{aligned}\Lambda_x &: P_{\mathcal{S}}[\varphi_x] \mapsto \mathrm{tr}_{\mathcal{W}+\mathcal{R}}[U_x(\rho_{\mathcal{W}} \otimes P_{\mathcal{S}}[\varphi_x] \otimes \tau^{\beta}_{\mathcal{R}})U_x^{\dagger}],\\ \Lambda'_x &: \tau^{\beta}_{\mathcal{R}} \mapsto \mathrm{tr}_{\mathcal{W}+\mathcal{S}}[U_x(\rho_{\mathcal{W}} \otimes P_{\mathcal{S}}[\varphi_x] \otimes \tau^{\beta}_{\mathcal{R}})U_x^{\dagger}],\\ \Lambda^*_x &: \rho_{\mathcal{W}} \mapsto \mathrm{tr}_{\mathcal{S}+\mathcal{R}}[U_x(\rho_{\mathcal{W}} \otimes P_{\mathcal{S}}[\varphi_x] \otimes \tau^{\beta}_{\mathcal{R}})U_x^{\dagger}].\end{aligned} \tag{C.2}$$

The subadditivity of the von-Neumann entropy and its invariance under unitary evolution implies that

$$S(\Lambda'_x[\tau^{\beta}_{\mathcal{R}}]) - S(\tau^{\beta}_{\mathcal{R}}) \geqslant S(\rho_{\mathcal{W}}) - S(\Lambda^*_x[\rho_{\mathcal{W}}]) - S(\Lambda_x[\varphi_x]). \tag{C.3}$$

Recalling that when feature 2 is satisfied, $S(\rho_{\mathcal{W}}) - S(\Lambda^*_x[\rho_{\mathcal{W}}]) = 0$, then by definition 2 and equation (C.3), the work that can be extracted for each measurement outcome, when both features 1 and 2 are satisfied, is bounded by

$$\begin{aligned}W_x &= \mathrm{tr}[H_{\mathcal{R}}(\tau^{\beta}_{\mathcal{R}} - \Lambda'_x[\tau^{\beta}_{\mathcal{R}}])] + \mathrm{tr}[H_{\mathcal{S}}(P_{\mathcal{S}}[\varphi_x] - \Lambda_x[\varphi_x])],\\ &= \beta^{-1}(S(\tau^{\beta}_{\mathcal{R}}) - S(\Lambda'_x[\tau^{\beta}_{\mathcal{R}}]) - S(\Lambda'_x[\tau^{\beta}_{\mathcal{R}}]\|\tau^{\beta}_{\mathcal{R}}))\\ &\quad + \mathrm{tr}[H_{\mathcal{S}}(P_{\mathcal{S}}[\varphi_x] - \Lambda_x[\varphi_x])],\\ &\leqslant \beta^{-1}(S(\Lambda_x[\varphi_x]) - S(\Lambda'_x[\tau^{\beta}_{\mathcal{R}}]\|\tau^{\beta}_{\mathcal{R}}))\\ &\quad + \mathrm{tr}[H_{\mathcal{S}}(P_{\mathcal{S}}[\varphi_x] - \Lambda_x[\varphi_x])],\\ &\leqslant \beta^{-1}S(\Lambda_x[\varphi_x]) + \mathrm{tr}[H_{\mathcal{S}}(P_{\mathcal{S}}[\varphi_x] - \Lambda_x[\varphi_x])].\end{aligned}$$

The final inequality can be saturated when the relative entropy term, $S(\Lambda'_x[\tau^{\beta}_{\mathcal{R}}]\|\tau^{\beta}_{\mathcal{R}})$, which is a non-negative number, is made vanishingly small. As shown in [12], this can be done if the dimension of $\mathcal{H}_{\mathcal{R}}$ is chosen to be sufficiently large, and its Hamiltonian spectrum is carefully chosen. As $S(\Lambda_x[\varphi_x])$ can be positive even when the weight's entropy is not allowed to change, we can always have positive work extraction. This is true even if the post-measurement state $|\varphi_x\rangle$ is the groundstate of $H_{\mathcal{S}}$. Moreover, if $H_{\mathcal{S}}$ is fully degenerate, and $\Lambda_x[\varphi_x] = \mathbb{1}_{\mathcal{S}}/d$, then the maximum value of W_x will be $K_{\mathrm{B}}T\ln(d)$ for all $x \in \mathcal{X}$. If $d = 2$, this coincides with the work extracted from the classical Szilard engine when the volumes of the left and right side of the partition are identical.

C.2. Degenerate observables

Recall that theorem 1 states that, when feature 2 is satisfied, then the extracted work will not be positive for the outcome where the post-measurement state coincides with the groundstate of the system Hamiltonian. Here we show that, if the observable M is both degenerate and is measured 'inefficiently', then the post-measurement states can always be chosen so as to have more energy than the groundstate of $H_{\mathcal{S}}$, thus allowing for the circumvention of theorem 1.

For a system $\mathcal{S}$ with Hilbert space $\mathcal{H}_{\mathcal{S}} \simeq \mathbb{C}^d$ such that $d > 2$, let $M_{\mathcal{S}}$ be a degenerate observable

$$M_{\mathcal{S}} = \sum_{x\in\mathcal{X}} xP^x_{\mathcal{S}}, \tag{C.4}$$

such that $|\mathcal{X}| < d$, and $\{P^x_{\mathcal{S}}\}_{x\in\mathcal{X}}$ is a complete and orthogonal set of projection operators on $\mathcal{H}_{\mathcal{S}}$. We label the orthonormal eigenstates of $M_{\mathcal{S}}$ as $|\varphi^{\alpha}_x\rangle$, where α is a degeneracy label, such that $M_{\mathcal{S}}|\varphi^{\alpha}_x\rangle = x|\varphi^{\alpha}_x\rangle$ for all α and x. The measurement model for this observable, $\mathscr{M} = (\mathcal{H}_{\mathcal{D}}, |\psi\rangle, U_M, Z_{\mathcal{D}})$, will be repeatable if for all $x \in \mathcal{X}$, the post-measurement states lie in the support of $P^x_{\mathcal{S}}$. Moreover, by the WAY theorem, if $\mathscr{M}$ is to be repeatable, given that U_M conserves the total Hamiltonian, then $P^x_{\mathcal{S}}$ must commute with $H_{\mathcal{S}}$ for all $x \in \mathcal{X}$. Consider the projector $P^y_{\mathcal{S}}$ whose support contains the groundstate(s) of $H_{\mathcal{S}}$. It follows that for a repeatable measurement, y is the only outcome whose post-measurement state will have support on the groundstate(s) of $H_{\mathcal{S}}$. Therefore, in order to circumvent theorem 1 we need to show that, for all $\rho_{\mathcal{S}}$, the post-measurement state given outcome y has more energy than the minimum eigenvalue of $H_{\mathcal{S}}$.

We will now look at two repeatable, and energy conserving measurement models for the degenerate observable $M_{\mathcal{S}}$. The first model is a generalization of a Lüders measurement [51, 52]. Here, for some state $\rho_{\mathcal{S}}$, the post-measurement state of outcome y is the groundstate of $H_{\mathcal{S}}$. Consequently, this measurement model will not circumvent theorem 1. In the second model, we may always ensure that the post-measurement state for outcome y will have more energy than the groundstate, thus circumventing theorem 1. We show that this is equivalent to coarse-graining the measurement outcomes of a non-degenerate observable, in such a way so as to allow for a repeatable measurement that is also 'inefficient'.

C.2.1. Strong value-correlation measurements. These measurements, just as the standard measurements for non-degenerate observables, have the property that, for any pure state $|\Psi\rangle \in \mathcal{H}_{\mathcal{S}}$, the post-measurement state for outcome $x \in \mathcal{X}$ will also be pure. Here, the premeasurement unitary operator is

$$U_M : |\varphi^{\alpha}_x\rangle \otimes |\psi\rangle \mapsto |\tilde{\varphi}^{\alpha}_x\rangle \otimes |\psi_x\rangle, \tag{C.5}$$

where $\{|\tilde{\varphi}^{\alpha}_x\rangle\}_{\alpha}$ is an orthonormal basis that spans $P^x_{\mathcal{S}}(\mathcal{H}_{\mathcal{S}})$. The instrument implemented by this measurement model will be

$$\mathcal{I}_x^M : \rho_{\mathcal{S}} \mapsto V_x P_{\mathcal{S}}^x \rho_{\mathcal{S}} P_{\mathcal{S}}^x V_x^\dagger, \tag{C.6}$$

where V_x is a unitary operator acting on the support of $P_{\mathcal{S}}^x$. This instrument has only one Kraus operator, $K_x = V_x P_{\mathcal{S}}^x$, and it is said to result in an 'efficient' measurement. If $V_x = \mathbb{1}$, whereby $|\tilde{\varphi}_x^\alpha\rangle = |\varphi_x^\alpha\rangle$, we have a Lüders measurement.

If the system is initially in the pure state

$$|\Psi\rangle = \sum_{x,\alpha} c_x^\alpha |\varphi_x^\alpha\rangle, \tag{C.7}$$

the post-measurement state for outcome y will be

$$\frac{\mathcal{I}_y^M(P_{\mathcal{S}}[\Psi])}{\mathrm{tr}[\mathcal{I}_y^M(P_{\mathcal{S}}[\Psi])]} = P_{\mathcal{S}}[\Psi_y],$$
$$|\Psi_y\rangle = \frac{1}{N}\sum_\alpha c_y^\alpha |\tilde{\varphi}_y^\alpha\rangle, \quad N^2 = \sum_\alpha |c_y^\alpha|^2. \tag{C.8}$$

Therefore, for some state $|\Psi\rangle$, the post-measurement state $|\Psi_y\rangle$ will be equal to the groundstate of the Hamiltonian. As such, theorem 1 will not be circumvented.

C.2.2. Coarse-grained standard measurements. Let us denote the degenerate eigenstates of $Z_{\mathcal{D}}$ as the orthonormal set of vectors $\{|\psi_x^\alpha\rangle\}$ such that $Z_{\mathcal{D}}|\psi_x^\alpha\rangle = x|\psi_x^\alpha\rangle$ for all x and α. The premeasurement unitary operator can then be defined as

$$U_M : |\varphi_x^\alpha\rangle \otimes |\psi\rangle \mapsto |\tilde{\varphi}_x^\alpha\rangle \otimes |\psi_x^\alpha\rangle. \tag{C.9}$$

Comparing with equation (2.2), we may see this as a coarse-grained measurement of a standard, non-degenerate observable. Now, the vectors in $\{|\tilde{\varphi}_x^\alpha\rangle\}_\alpha$ no longer have to be orthonormal. But, they must still be eigenstates of $M_{\mathcal{S}}$ with eigenvalue x for the measurement to be repeatable. The instrument implemented by this measurement model will be

$$\mathcal{I}_x^M : \rho_{\mathcal{S}} \mapsto \sum_\alpha V_{x,\alpha} P_{\mathcal{S}}[\varphi_x^\alpha] \rho_{\mathcal{S}} P_{\mathcal{S}}[\varphi_x^\alpha] V_{x,\alpha}^\dagger, \tag{C.10}$$

where $V_{x,\alpha}$ are unitary operators acting on the support of $P_{\mathcal{S}}^x$. In contrast to the generalized Lüders measurement discussed previously, this instrument has more than one Kraus operator, and leads to an 'inefficient' measurement.

If the system is initially in the pure state

$$|\Psi\rangle = \sum_{x,\alpha} c_x^\alpha |\varphi_x^\alpha\rangle, \tag{C.11}$$

the post-measurement state for outcome y will be

$$\frac{\mathcal{I}_y^M(P_{\mathcal{S}}[\Psi])}{\mathrm{tr}[\mathcal{I}_y^M(P_{\mathcal{S}}[\Psi])]} = \frac{1}{\sum_\alpha |c_y^\alpha|^2}\sum_\alpha |c_y^\alpha|^2 P_{\mathcal{S}}[\tilde{\varphi}_y^\alpha]. \tag{C.12}$$

Due to the orthogonality of the vectors $|\psi_x^\alpha\rangle$ in equation (C.9), for each x and α, the vectors $|\tilde{\varphi}_y^\alpha\rangle$ can be any superpositions of Hamiltonian eigenstates that live in the support of $P_{\mathcal{S}}^y$. So we may simply choose these as the highest energy state within that subspace. Consequently, theorem 1 will be circumvented.

Appendix D. Net work extraction per cycle of a quantum Szilard engine without heat from a thermal reservoir

Each cycle of work extraction involves the following steps: (i) $\mathcal{S}$ is given in state $\rho_{\mathcal{S}}$; (ii) $\mathcal{S}$ and $\mathcal{D}$ undergo a joint unitary evolution by U_M; (iii) work is extracted from $\mathcal{S}$ by a feedback unitary operator V on $\mathcal{W} + \mathcal{S} + \mathcal{D}$; (iv) $\mathcal{D}$ is reset to its initial state $|\psi\rangle$ by coupling to a thermal reservoir. Figure 2 shows this schematically.

The initial state of the compound $\mathcal{W} + \mathcal{S} + \mathcal{D} + \mathcal{R}$ is

$$\rho = \rho_{\mathcal{W}} \otimes \rho_{\mathcal{S}} \otimes P_{\mathcal{D}}[\psi] \otimes \tau_{\mathcal{R}}^\beta, \tag{D.1}$$

where $\tau_{\mathcal{R}}^\beta := \mathrm{e}^{-\beta H_{\mathcal{R}}}/\mathrm{tr}[\mathrm{e}^{-\beta H_{\mathcal{R}}}]$ is the Gibbs state of the reservoir at inverse temperature $\beta = (K_{\mathrm{B}}T)^{-1}$. After premeasurement, objectification, and feedback the state will be

$$\rho' := V(\rho_{\mathcal{W}} \otimes \rho_{\mathcal{S}+\mathcal{D}}^{M,\mathrm{O}})V^\dagger \otimes \tau_{\mathcal{R}}^\beta, \tag{D.2}$$

where $\rho^{M,O}_{\mathcal{S}+\mathcal{D}}$ is defined in equation (2.5). The marginal states of ρ' satisfy the relations

$$\begin{aligned}\rho'_{\mathcal{S}} &:= \sum_{x\in\mathcal{X}} p^{M}_{\rho_{\mathcal{S}}}(x)\Lambda_x[\tilde{\varphi}_x] \equiv \mathrm{tr}_{\mathcal{W}+\mathcal{D}}[V(\rho_{\mathcal{W}}\otimes\rho^{M,O}_{\mathcal{S}+\mathcal{D}})V^\dagger],\\ \rho'_{\mathcal{W}} &:= \sum_{x\in\mathcal{X}} p^{M}_{\rho_{\mathcal{S}}}(x)\Lambda^*_x[\rho_{\mathcal{W}}] \equiv \mathrm{tr}_{\mathcal{S}+\mathcal{D}}[V(\rho_{\mathcal{W}}\otimes\rho^{M,O}_{\mathcal{S}+\mathcal{D}})V^\dagger],\\ \rho'_{\mathcal{D}} &:= \mathrm{tr}_{\mathcal{W}+\mathcal{S}}[V(\rho_{\mathcal{W}}\otimes\rho^{M,O}_{\mathcal{S}+\mathcal{D}})V^\dagger],\end{aligned} \tag{D.3}$$

where $p^{M}_{\rho_{\mathcal{S}}}(x)$ is the Born rule probability defined in equation (2.6), while Λ_x and Λ^*_x are the CPTP maps induced by feedback, as defined in equation (2.19).

Using definition 2, we may view the work transferred into the weight, when the different measurement outcomes are not distinguished from one another, to be

$$\begin{aligned}W_{\mathcal{X}} &:= \mathrm{tr}[H_{\mathcal{W}}(\rho'_{\mathcal{W}}-\rho_{\mathcal{W}})] + K_{\mathrm{B}}T\,(S(\rho_{\mathcal{W}})-S(\rho'_{\mathcal{W}})),\\ &= \mathrm{tr}[H_{\mathcal{S}}(\rho_{\mathcal{S}}-\rho'_{\mathcal{S}})] + \mathrm{tr}[H_{\mathcal{D}}(P_{\mathcal{D}}[\psi]-\rho'_{\mathcal{D}})] + K_{\mathrm{B}}T\,(S(\rho_{\mathcal{W}})-S(\rho'_{\mathcal{W}})).\end{aligned} \tag{D.4}$$

Here we have used the fact that feedback and measurement are energy conserving on the total system. We call $W_{\mathcal{X}}$ the 'coarse-grained' work, which is different to the average work, obtained by averaging W_x over all measurement outcomes $x\in\mathcal{X}$, which is

$$\begin{aligned}\langle W_x\rangle &:= \sum_{x\in\mathcal{X}} p^{M}_{\rho_{\mathcal{S}}}(x)\,W_x,\\ &= \mathrm{tr}[H_{\mathcal{W}}(\rho'_{\mathcal{W}}-\rho_{\mathcal{W}})] + K_{\mathrm{B}}T\left(S(\rho_{\mathcal{W}}) - \sum_{x\in\mathcal{X}} p^{M}_{\rho_{\mathcal{S}}}(x)S(\Lambda^*_x[\rho_{\mathcal{W}}])\right),\\ &\geqslant W_{\mathcal{X}}.\end{aligned} \tag{D.5}$$

The inequality here is due to the concavity of the von-Neumann entropy.

Before the cycle can begin anew, the demon must be reset to the original pure state $|\psi\rangle$. This is achieved within the Landauer framework, by coupling $\mathcal{D}$ with $\mathcal{R}$ by the 'erasure' unitary operator $U_R : \mathcal{H}_{\mathcal{D}}\otimes\mathcal{H}_{\mathcal{R}} \to \mathcal{H}_{\mathcal{D}}\otimes\mathcal{H}_{\mathcal{R}}$. If the reservoir is infinitely large, then U_R can be chosen so that

$$\mathrm{tr}_{\mathcal{R}}[U_R(\rho'_{\mathcal{D}}\otimes\tau^{\beta}_{\mathcal{R}})U_R^\dagger] = P_{\mathcal{D}}[\psi]. \tag{D.6}$$

To be sure, U_R is generally not energy conserving, and thus needs a hidden work source. Notwithstanding, this is not a problem, because erasure always consumes work. Therefore, this hidden work source does not contribute to work extraction within a cycle. Defining the reduced state of the reservoir after its interaction with $\mathcal{D}$ as $\tau'_{\mathcal{R}}$, the consequent increase in energy of the reservoir, defined as heat, obeys Landauer's inequality

$$Q := \mathrm{tr}[H_{\mathcal{R}}(\tau'_{\mathcal{R}}-\tau^{\beta}_{\mathcal{R}})] \geqslant \beta^{-1}S(\rho'_{\mathcal{D}}). \tag{D.7}$$

As shown in [12], this bound can be achieved if the reservoir is infinitely large, and its Hamiltonian has a specific spectrum. Furthermore, we note that premeasurement, objectification, and feedback results in a unital CPTP map, which does not decrease the von-Neumann entropy [57, 58]. This, together with the subadditivity of the von-Neumann entropy [59], implies that

$$\begin{aligned}S(\rho_{\mathcal{W}})+S(\rho_{\mathcal{S}}) &= S(\rho_{\mathcal{W}}\otimes\rho_{\mathcal{S}}\otimes P_{\mathcal{D}}[\psi]),\\ &\leqslant S(V(\rho_{\mathcal{W}}\otimes\rho^{M,O}_{\mathcal{S}+\mathcal{D}})V^\dagger),\\ &\leqslant S(\rho'_{\mathcal{W}})+S(\rho'_{\mathcal{S}})+S(\rho'_{\mathcal{D}}).\end{aligned} \tag{D.8}$$

Consequently, by combining equations (D.7) and (D.8), and also taking into account the energy change of the demon due to erasure, the work cost of erasure is shown to obey the inequality

$$\begin{aligned}W_R &:= \mathrm{tr}[H_{\mathcal{D}}(P_{\mathcal{D}}[\psi]-\rho'_{\mathcal{D}})] + Q,\\ &\geqslant \mathrm{tr}[H_{\mathcal{D}}(P_{\mathcal{D}}[\psi]-\rho'_{\mathcal{D}})] + K_{\mathrm{B}}T\,(S(\rho_{\mathcal{W}})+S(\rho_{\mathcal{S}})-S(\rho'_{\mathcal{W}})-S(\rho'_{\mathcal{S}})).\end{aligned} \tag{D.9}$$

Defining the net coarse-grained work extraction as $W^{\mathrm{net}}_{\mathcal{X}} := W_{\mathcal{X}} - W_R$, by combining equations (D.4) and (D.9) we arrive at the inequality

$$\begin{aligned}W^{\mathrm{net}}_{\mathcal{X}} &= F(\rho'_{\mathcal{W}}) - F(\rho_{\mathcal{W}}) - W_R = \mathrm{tr}[H_{\mathcal{S}}(\rho_{\mathcal{S}}-\rho'_{\mathcal{S}})]\\ &\quad + K_{\mathrm{B}}T\,(S(\rho_{\mathcal{W}})-S(\rho'_{\mathcal{W}})) - Q, \leqslant F(\rho_{\mathcal{S}}) - F(\rho'_{\mathcal{S}}).\end{aligned} \tag{D.10}$$

The net average work extraction $\langle W^{\mathrm{net}}_x\rangle := \langle W_x\rangle - W_R$, on the other hand, obeys the modified inequality

$$\langle W^{\mathrm{net}}_x\rangle \leqslant F(\rho_{\mathcal{S}}) - F(\rho'_{\mathcal{S}}) + K_{\mathrm{B}}T\left(S(\rho'_{\mathcal{W}}) - \sum_{x\in\mathcal{X}} p^{M}_{\rho_{\mathcal{S}}}(x)S(\Lambda^*_x[\rho_{\mathcal{W}}])\right). \tag{D.11}$$

Therefore, we see that while the coarse-grained work definition of equation (D.4) will satisfy the second law, the average work extraction defined in equation (D.5) will not; if $\rho_{\mathcal{S}}$ is initially thermal, the net coarse-grained work

extraction given by equation (D.10) will never be positive, whereas the net average work extraction given by equation (D.11) could be.

ORCID iDs

M Hamed Mohammady https://orcid.org/0000-0002-0443-5242

References

[1] Sagawa T and Ueda M 2012 *Phys. Rev.* E **85** 021104
[2] Shiraishi N, Ito S, Kawaguchi K and Sagawa T 2015 *New J. Phys.* **17** 045012
[3] Maxwell J C 1871 *Theory of Heat* (London: Longmans)
[4] Maruyama K, Nori F and Vedral V 2009 *Rev. Mod. Phys.* **81** 1
[5] Szilard L 1929 *Z. Phys.* **53** 840
[6] Balian R 2007 *From Microphysics to Macrophysics* vol 1 (Berlin: Springer)
[7] Penrose O 1970 *Foundations of Statistical Mechanics: A Deductive Treatment* (Oxford: Pergamon)
[8] Bennett C H 1982 *Int. J. Theor. Phys.* **21** 905
[9] Bennett C H 2003 *Stud. Hist. Phil. Mod. Phys.* **34** 501
[10] Landauer R 1961 *IBM J. Res. Dev.* **5** 183
[11] Landauer R 1996 *Phys. Lett.* A **217** 188
[12] Reeb D and Wolf M M 2014 *New J. Phys.* **16** 103011
[13] Vinjanampathy S and Anders J 2016 *Contemp. Phys.* **57** 545
[14] Goold J, Huber M, Riera A, del Rio L and Skrzypczyk P 2016 *J. Phys. A: Math. Theor.* **49** 143001
[15] Millen J and Xuereb A 2016 *New J. Phys.* **18** 011002
[16] Horodecki M and Oppenheim J 2013 *Nat. Commun.* **4** 2059
[17] Kammerlander P and Anders J 2015 *Sci. Rep.* **6** 22174
[18] Lostaglio M, Jennings D and Rudolph T 2015 *Nat. Commun.* **6** 6383
[19] Perarnau-Llobet M, Hovhannisyan K V, Huber M, Skrzypczyk P, Brunner N and Acín A 2015 *Phys. Rev.* X **5** 041011
[20] Gogolin C and Eisert J 2016 *Rep. Prog. Phys.* (https://doi.org/10.1088/0034-4885/79/5/056001)
[21] Guryanova Y, Popescu S, Short A J, Silva R and Skrzypczyk P 2016 *Nat. Commun.* **7** 12049
[22] Yunger Halpern N, Faist P, Oppenheim J and Winter A 2016 *Nat. Commun.* **7** 12051
[23] Alhambra Á M, Masanes L, Oppenheim J and Perry C 2016 *Phys. Rev.* X **6** 041017
[24] Zurek W H 1986 Maxwell's demon, Szilard's engine and quantum measurements *Frontiers of Nonequilibrium Statistical Physics* ed G T Moore and M O Scully (Boston, MA: Springer) pp 151–61
[25] Plesch M, Dahlsten O, Goold J and Vedral V 2014 *Sci. Rep.* **4** 6995
[26] Kim S W, Sagawa T, De Liberato S and Ueda M 2011 *Phys. Rev. Lett.* **106** 070401
[27] Sagawa T and Ueda M 2008 *Phys. Rev. Lett.* **100** 080403
[28] Jacobs K 2009 *Phys. Rev.* A **80** 012322
[29] Park J J, Kim K-H, Sagawa T and Kim S W 2013 *Phys. Rev. Lett.* **111** 230402
[30] Camati P A, Peterson J P S, Batalhão T B, Micadei K, Souza A M, Sarthour R S, Oliveira I S and Serra R M 2016 *Phys. Rev. Lett.* **117** 240502
[31] Cottet N *et al* 2017 *PNAS* **114** 7561
[32] Elouard C, Herrera-Martí D, Clusel M and Auffèves A 2017 *npj Quantum Inf.* **3** 9
[33] Elouard C, Herrera-Martí D, Huard B and Auffèves A 2017 *Phys. Rev. Lett.* **118** 260603
[34] Sagawa T and Ueda M 2009 *Phys. Rev. Lett.* **102** 250602
[35] Jacobs K 2012 *Phys. Rev.* E **86** 040106
[36] Navascués M and Popescu S 2014 *Phys. Rev. Lett.* **112** 140502
[37] Miyadera T 2016 *Found. Phys.* **46** 1522
[38] Abdelkhalek K, Nakata Y and Reeb D 2016 arXiv:1609.06981
[39] Wigner E 1952 *Z. Phys.* **133** 101
[40] Araki H and Yanase M M 1960 *Phys. Rev.* **120** 622
[41] Miyadera T and Imai H 2006 *Phys. Rev.* A **74** 024101
[42] Loveridge L and Busch P 2011 *Eur. Phys. J.* D **62** 297
[43] Busch P and Loveridge L 2011 *Phys. Rev. Lett.* **106** 110406
[44] Ahmadi M, Jennings D and Rudolph T 2013 *New J. Phys.* **15** 013057
[45] Skrzypczyk P, Short A J and Popescu S 2014 *Nat. Commun.* **5** 4185
[46] Åberg J 2014 *Phys. Rev. Lett.* **113** 150402
[47] von Neumann J 1996 *Mathematical Foundations of Quantum Mechanics* (Princeton, NJ: Princeton University Press)
[48] Busch P, Grabowski M and Lahti P J 1995 *Operational Quantum Physics* (Berlin: Springer)
[49] Busch P, Lahti P J and Mittelstaedt P 1996 *The Quantum Theory of Measurement* (Berlin: Springer)
[50] Busch P, Lahti P J, Pellonpää J P and Ylinen K 2016 *Quantum Measurement* (Berlin: Springer)
[51] Heinosaari T and Ziman M 2011 *The Mathematical Language of Quantum Theory* (Cambridge: Cambridge University Press)
[52] Mittelstaedt P 2004 *The Interpretation of Quantum Mechanics and the Measurement Process* (Cambridge: Cambridge University Press)
[53] Gemmer J and Anders J 2015 *New J. Phys.* **17** 085006
[54] Gallego R, Eisert J and Wilming H 2016 *New J. Phys.* **18** 103017
[55] Morikuni Y, Tajima H and Hatano N 2017 *Physi. Rev.* E **95** 032147
[56] Anders J and Giovannetti V 2013 *New J. Phys.* **15** 033022
[57] Alberti P and Uhlmann A 1982 *Stochasticity and Partial Order: Doubly Stochastic Maps and Unitary Mixing* (Berlin: Springer)
[58] Nakahara M, Rahimi R and Saitoh A 2008 *Decoherence Suppression in Quantum Systems* (Singapore: World Scientific)
[59] Petz D 2008 *Quantum Information Theory and Quantum Statistics* (Berlin: Springer)